International Journal of Medical Informatics 128 (2019) 62–70

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer

Yixuan Tang[a], Jisong Yang[a], Pei San Ang[b], Sreemanee Raaj Dorajoo[b], Belinda Foo[b], Sally Soh[b], Siew Har Tan[b], Mun Yee Tham[b], Qing Ye[b,c], Lynette Shek[d], Cynthia Sung[b,e], Anthony Tung[a,*]

[a] Department of Computer Science, School of Computing, National University of Singapore, Singapore
[b] Vigilance & Compliance Branch, Health Products Regulation Group, Health Sciences Authority, Singapore
[c] Genome Institute of Singapore, Agency for Science and Technology, Singapore
[d] Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore
[e] Health Services and Systems Research, Duke-NUS Medical School, Singapore

ARTICLE INFO

Keywords:
Pharmacovigilance
Text mining
Electronic medical records
Expert system
Adverse drug reaction

ABSTRACT

Background: Hospital discharge summaries offer a potentially rich resource to enhance pharmacovigilance efforts to evaluate drug safety in real-world clinical practice. However, it is infeasible for experts to read through all discharge summaries to find cases of drug-adverse event (AE) relations.

Purpose: The objective of this paper is to develop a natural language processing (NLP) framework to detect drug-AE relations from unstructured hospital discharge summaries.

Basic procedures: An NLP algorithm was designed using customized dictionaries of drugs, adverse event (AE) terms, and rules based on trigger phrases, negations, fuzzy logic and word distances to recognize drug, AE terms and to detect drug-AE relations. Furthermore, a customized annotation tool was developed to facilitate expert review of discharge summaries from a tertiary hospital in Singapore in 2011.

Main findings: A total of 33 trial sets with 50 to 100 records per set were evaluated (1620 discharge summaries) by our algorithm and reviewed by pharmacovigilance experts. After every 6 trial sets, drug and AE dictionaries were updated, and rules were modified to improve the system. Excellent performance was achieved for drug and AE entity recognition with over 92% precision and recall. On the final 6 sets of discharge summaries (600 records), our algorithm achieved 75% precision and 59% recall for identification of valid drug-AE relations.

Principal conclusions: Adverse drug reactions are a significant contributor to health care costs and utilization. Our algorithm is not restricted to particular drugs, drug classes or specific medical specialties, which is an important attribute for a national regulatory authority to carry out comprehensive safety monitoring of drug products. Drug and AE dictionaries may be updated periodically to ensure that the tool remains relevant for performing surveillance activities. The development of the algorithm, and the ease of reviewing and correcting the results of the algorithm as part of an iterative machine learning process, is an important step towards use of hospital discharge summaries for an active pharmacovigilance program.

1. Introduction

The growing availability of electronic medical records opens up opportunities for more comprehensive capturing of adverse drug reactions (ADR) using mining algorithms. For drug regulatory authorities who have a public health mission to ensure drug safety for a national population, this is important because clinical trials typically investigated a few thousands of patients who are enrolled under strict and extensive inclusion and exclusion criteria, such as age and comorbidity restrictions, which may not be representative of the general population. The true extent and types of ADRs may only come to light in the post-market phase when the drug is used in the general population with a broader variety of demographic characteristics and health states. Pharmacovigilance programs rely on adverse event (AE) reporting from companies and physicians, and some accept reporting from patients. However, these programs are known to have under-reporting bias, and it is estimated that only 2% to 18% of all AEs are actually captured by current reporting systems [1]. Therefore, efficient methods to fully capture the extent and types of drug-AE relations are needed. In order to gain a more comprehensive view of the overall ADR landscape in

* Corresponding author.
E-mail address: anthony@comp.nus.edu.sg (A. Tung).

https://doi.org/10.1016/j.ijmedinf.2019.04.017
Received 2 November 2018; Received in revised form 22 March 2019; Accepted 21 April 2019

Singapore, one approach we are exploring is text mining of hospital discharge summaries.

Hospital discharge summaries are free text narratives that summarize patients medical conditions, drug allergies, AEs, laboratory investigations, procedures, medications and outcomes during hospitalization. They can serve as a good resource for carrying out pharmacovigilance activities. However, having experts to read through all the discharge summaries to find cases of drug-AE relations is infeasible. To develop an annotation tool to mine drug-AE relations automatically from free text hospital discharge summaries will be challenging due to the variety of contents, structures, styles, abbreviations, spelling errors, acronyms, sentence fragments and ungrammatical constructs appearing in the discharge summaries [2].

There have been several studies focusing on extracting drug-AE relations from unstructured electric health records. Existing approaches can be divided into three main categories, namely statistics-based methods, keyword-based methods and learning-based methods. Statistics-based methods calculate the co-occurrence of a pair of drug and AE based on a large corpora of medical reports and then detect potential drug-AE pairs from an overall view. These studies had not aimed to pick up semantically connected drug-AE pairs but rather identify frequency co-occurrences of drugs and AE names that are present in the same discharge summary [1,3].

Keyword-based methods identify a collection of trigger phrases which may indicate the presence of drug-AE relations and then employ customized rules on the syntactic and semantic patterns surrounding the drugs and AE to further determine the relation. Ballard et al. adopted a text search algorithm to recognize surgical device related adverse events for urogynecologic mesh [4]. Ferrajolo et al. conducted a retrospective EHR-based cohort study to investigate potential drug-induced acute liver injury [5]. Pathak et al. applied a trigger phrase based search within queries of resource description framework (RDF) graphs to detect drug-AE relations for cardiovascular and gastroenterology drugs [6]. Eriksson et al. aimed to identify drug-AE pairs in Danish clinical narratives. They first applied text matching to identify AEs based on an AE dictionary constructed in-house and then applied post-processing rules to handle negations and to connect AE to drugs [2]. Kilbridge et al. designed an expert system to flag certain combinations of structured laboratory and pharmacy data to identify records with possible AEs from a pediatric population, and these were then sent for expert review [7]. Miura et al. adopted dependency parsing to extract sentence features to predict whether a drug-AE relationship exists within a single sentence [8]. In all, these studies either target specific disease related AEs or limit the distance between drugs and AEs to be within one sentence.

There are some learning-based methods for classifying drug-AE relations. Visweswaran et al. investigated four naive Bayes models to calculate the probability that a discharge summary contains an AE, given the existence of certain medical terms in the document, although the method does not extract the semantically associated drug-AE pairs in the discharge summary [9]. Sohn et al. combined rules and C4.5 classifiers to extract drug-AE relations from psychiatric clinical notes [10]. More recently, Munkhdalai et al. investigated the performance of model-based systems to identify relations (including adverse reactions) in free-text EHR documents and report the performance of three learning model, namely Support Vector Machine (SVM), Recurrent Neural Network-based long short-term memory (LSTM) models and a supervised descriptive rule induction. They used a dataset of 971 annotated EHR texts and report that SVM model outperforms the other methods tested [11]. Conversely, Santiso et al. employed the use of a Joint AB-LSTM model incorporating context-aware architectures and compared it against a Random Forest model to detect true drug-adverse event pairs after developing word embeddings on a large unannotated corpus. They had used annotated text documents from 2 sources, amounting to a total of 336 documents. Santiso et al. report that the Joint AB-LSTM model outperforms the Random Forest model by approximately 16% on their testing dataset [12,13]. Also, in a recent drug-AE relation recognition competition named “MADE1.0”, all the top performing teams adopted deep learning models which consist of an encoding layer, several bi-LSTM layers and conditional random field (CRF) layers for drug-AE relation identification [14].

In this paper, we employ a different approach to previous studies in this area which use computational learning models. We aimed to develop expert-derived rules to detect the presence of drug-AE pairs within inpatient discharge summaries. This is an extension of our previous work of developing a rule-based method to automatically extract drugs and AEs from hospital discharge summaries [15]. We describe an iterative approach in extracting drug-AE relations and to evaluate the performance of the rule-based method to identify drug-AE relations from unstructured discharge summaries.

2. Methodology

De-identified hospital discharge summaries were obtained from the National University Hospital (NUH) in Singapore for the period January 2011 to December 2011 after approval from the Domain Specific Review Board. NUH is a tertiary hospital with a full suite of medical, surgical and dental specialities. We curated a reference dataset with annotations indicating the presence of true drug-AE pairs. A total of 1800 discharge summaries were randomly selected to capture the representative sample of discharge summaries from all inpatient specialities. Only discharge summaries that were deemed complete and of substantial content were included (at least 3KB in size per text file containing one discharge summary). The algorithm developed was then applied to a subset of the discharge summaries and the results were evaluated by domain experts. Based on the results, the algorithm was updated iteratively and retested on a new set of discharge summaries.

We developed a rule-based approach named Readpeer for Active Pharmacovigilance (REAP), a customized annotation tool to address the problem. As illustrated in Fig. 1, the overall framework can be divided into two steps, named entity recognition (NER) and drug-AE relation extraction. Given a discharge summary, REAP starts with three common NLP pre-processing steps, namely tokenization, sentence detection and lemmatization. The drug names and AE names were identified from the discharge summary based on the dictionaries. The NER step is aided by fuzzy logic for misspellings and negation detection algorithms. Based on the named entities recognized, candidate drug-AE pairs are then generated and extracted if they meet the pre-defined distance criteria and the rules for trigger phrases. Finally, machine annotated results of REAP were evaluated by human experts using Readpeer. Each step is described in detail as below.

2.1. Named entity recognition

Though there are a few NER tools for medical terminologies available, such as MetaMap [16] and cTAKEs [17], we choose to develop the algorithm for drug and AE names extraction for two main reasons. Firstly, existing tools do not provide spell checking functionalities and do not handle abbreviations, which are very commonly used by doctors in free text discharge summaries. Secondly, existing tools are not flexible enough to facilitate customization of drug and AE dictionaries. Therefore, we developed an in-house approach to extract drug and AE names with fuzzy logic, negation detection and a list of customized dictionaries, including drug names, AE names, common English words, negation terms, connecting words and phrases (e.g. “secondary to”, “following administration of”, “due to”), drug-AE pairs and words related to vaccines [15].

The drug gazetteer contained a wide range of words and phrases that indicate drugs in discharge summaries. Reference sources used in curating the drug gazetteer included the registered drug product list maintained by the Health Sciences Authority, Singapore (HSA), brand names registered in other countries, RxNorm, FDA Orange Book and

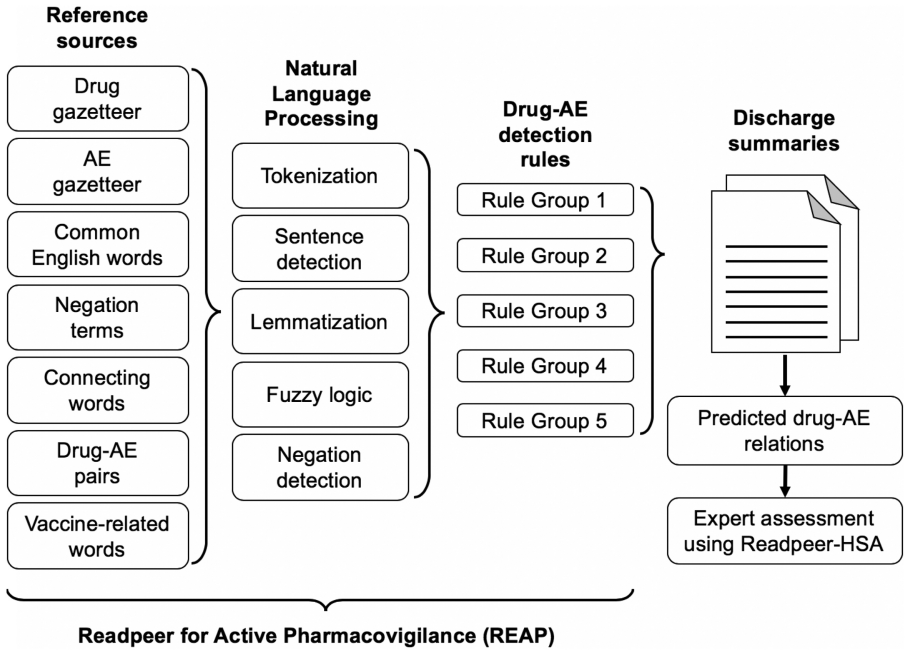

Fig. 1. System overview.

commonly used drug classes and abbreviations. The list was further expanded by common misspellings of drug names, which were identified when cleaning free text prescriptions from the NUH pharmacy database. The AE gazetteer was built based on the collection of the lowest level terms from the World Health Organization Adverse Reaction Terminology (WHO-ART) and the Medical Dictionary for Regulatory Activities (MedDRA). Four AE-unrelated system organ classes were removed from MedDRA, including investigations, injury poisoning and procedural complications, social circumstances, and surgical and medical procedures. Also, both gazetteers were modified with insertions or deletions through iterative reviews of machine annotated results by human experts during rule refinement. All annotators possessed a minimum qualification of a Bachelor degree in Pharmacy, in addition to prior experience working in hospitals and reviewing spontaneous AE reports submitted to HSA. At the time of this report, the drug gazetteer contained 8973 unique drug name phrases or abbreviations and 103,531 common misspellings, while the AE gazetteer contained 62,912 AE name phrases or abbreviations.

With the gazetteers in place, tagging of drug and AE names in the discharge summaries were carried out using text searching methods. The discharge summaries were first tokenized, split into sentences and transferred into lemma forms using Stanford CoreNLP tools. To check if a word or phrase appearing in a discharge summary was in the drug or AE gazetteers, data structures such as the inverted index and hash tables were used. A hash table allows an algorithm to check whether a word is present in the gazetteer within constant time, while inverted index helps to locate the phrases containing the word given. With these techniques, REAP was able to extract drug and AE names from discharge summaries at a speed of 120 summaries per minute.

2.1.1. Fuzzy logic

Spelling errors are a common feature in free text discharge summaries. Although many misspellings of drug and AE names were found during the training phase and had been incorporated into the gazetteers, several unmatched terms with spelling errors remained. To tackle those spelling errors, two fuzzy logic strategies were employed. Firstly, we identified words which could be matched to gazetteers within one character modification, including a single addition, deletion or substitution. To decrease false positives, this strategy was only applied on long words containing seven or more characters. Simultaneously, we applied Soundex, a phonetic algorithm, to index words by sound so that homophones were encoded to the same representation. Words or phrases which matched the sound of phrases in the gazetteers were recognized as drug or AE names even if the Levenshtein distance was larger than one. By introducing Soundex, British and American spellings such as anaemia-anemia, haemmorrhage-hemorrhage could be successfully recognized, as long as one of the pairs was included in our gazetteer. As a further enhancement, fuzzy logic was turned off on the 9000 most frequent English words in Google N-gram to prevent common English words from being detected as named entities.

2.1.2. Negation detection

To avoid tagging drugs which were not given to patients or AEs which appeared in text but did not specifically refer to the patient experiencing them, REAP used a list of negation phrases as a guide to avoid tagging of these terms even though they appear in the text. After identifying negation phrases, NegEx [18], a popular NLP algorithm for negation detection, was used to determine the negation scope within five words. However, it was not able to detect negated phrases outside five words. For example, in the following phrase, “no new established territorial infarct or intracranial haemorrhage observe”, NegEx failed to detect “intracranial haemorrhage” as negated. To remedy this problem, the following regular expression was used to determine negation scope.

$$\text{negation\_p hrase(word\{1, 5\}(|negation\_p hrase|and|or|,)?)\{1, 4\}} \quad (1)$$

The regular expression started with a negation phrase such as “no”, followed by a pattern that could occur one to four times. The pattern could have one to five words and an optional negation phrase, connected by “and”, “or” or “,”. Notably a full-stop “.” was not considered to be a word in order to ensure that the regular expression would not negate outside a sentence. For instance, “no territorial infarct. Intracranial haemorrhage exists”, “Intracranial haemorrhage” would not be negated because of the presence of a full stop in front. To avoid negating cases like “vomit” in “No fever but vomit continues”’, the phrase that followed “but” was not negated.

2.2. Rule-based relation extraction

After drug and AE names were extracted, candidate drug-AE pairs were generated. Previous publications [8,9] had assumed that the drug

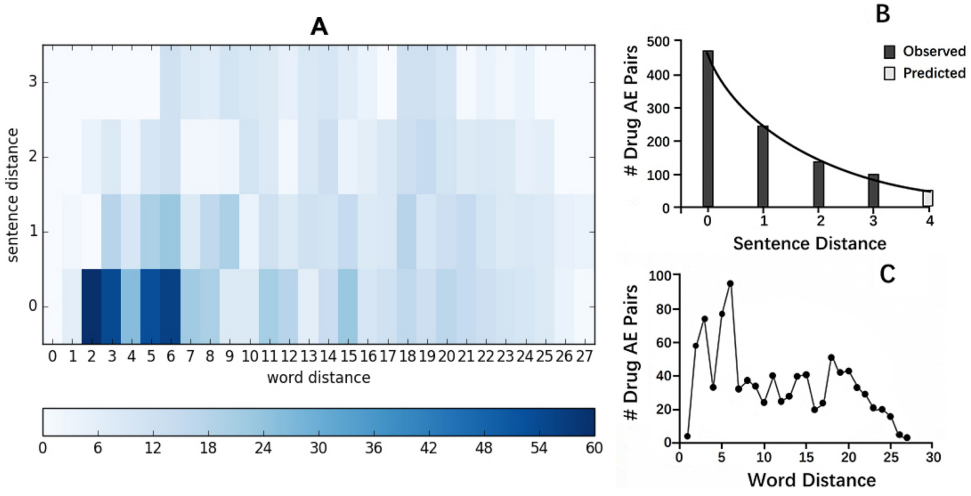

Fig. 2. Distribution of drug-AE pairs.

and AE for a drug-AE relation must reside in the same sentence. However, during initial reviews of discharge summaries, we found that many drugs and AEs for drug-AE relations were located in different sentences. To evaluate the sentence distances between the drug and AE for drug-AE relations, we collected statistics from a random selection of 1000 discharge summaries from the training data and the sentence and word distances for all the drug-AE pairs were visually inspected.

Fig. 2 shows the distribution of sentence distance and word distance between drug and AE names of 950 true drug-AE relations from 1000 discharge summaries. Sub-figure A is a heatmap providing an overview of how drug-AE pairs distribute along sentence distance and word distance. Though related drug and AE terms typically appear within 6 words in the same sentence, sometimes drug or AE terms may appear in different sentences or far away in terms of word distance. Sub-figures demonstrate the distribution of drug-AE pairs along each dimension more clearly. Sub-figure B illustrates that the sentence distance between drug and AE terms decreases considerably between a distance of 0 (i.e. both terms appearing in the same sentence) to 3 sentences apart. By extrapolation, we estimated that setting a sentence boundary of 3 sentences would capture the vast majority of drug-AE pairs. Having noticed that punctuation marks were generally used sparingly in some discharge summaries, a word boundary was also necessary for candidate pair generation as a supplementary restriction to sentence distance. As shown in sub-figure C, while drug-AE pairs were most frequently found between 2 to 7 words, the word distance of 8 to 25 also covered a large portion of true pairs. Based on the visual inspection of these plots, we decided to restrict candidate drug-AE pair searches to within 3 sentences or 25 words, because it captures most instances of valid drug-AE pairs while pruning the rest drug-AE pairs for efficiency.

$$\text{word\_d is}(\text{drug, AE}) < 25 \wedge \text{sentence\_d is}(\text{drug, AE}) < 3 \quad (2)$$

Likewise, the allowable distance between drug terms and trigger phrases as well as between AE terms and trigger phrases were capped at 25 words.

$$\text{word\_d is}(\text{drug, trigger\_p hrase}) < 25 \wedge \text{word\_d is}(\text{AE, trigger\_p hrase}) < 25 \quad (3)$$

After candidate drug-AE pairs within the allowable distances were identified, a more specific set of rules were applied, based on specific trigger phrase connecting the drug and AE terms in question. A total of 67 trigger phrases as listed in Appendix A were used. As shown in Table 1, representations of true drug-AE pairs such as ‘drug Cause AE’ (Rule 1) or ‘AE AttributeTo drug’ (Rule 2) were developed to capture a potential connection between the drug and AE as written by the treating physician. The italicized words represented a set of phrases of similar meaning or purpose in that context to delineate a drug-AE relation, and examples of that pattern are also provided in Table 1. Drug allergies were typically written as AllergyTo followed by the drug name (Rule 3). Other words or phrases indicated that the doctor took action to stop a drug or change a drug, and the typical pattern was ‘drug StopAfter AE’ or ‘StopBefore drug’ (Rule 4). Another pattern had the structure ‘AEBefore AE’ or ‘AE AEAfter’ (Rule 5). Drug-AE pairs that satisfied at least one specific rule were annotated for human evaluation.

2.3. Readpeer system

Readpeer-HSA is an annotation platform that we customized to facilitate expert evaluation of machine annotated results. As shown in Fig. 3, the discharge summary is displayed on the left while machine annotations are displayed as annotations on the right. In the discharge summary, drugs are highlighted in blue and AEs are highlighted in green. When the mouse hovers over a drug-AE relation on the right, the

Table 1
Rule group examples.

No	Relation rule	Phrase set	Example
1	drug *Cause* AE	*Cause*: {caused, induced, held off in in view of, resulted in, …}	Isoniazid induced DILI
2	AE *AttributeTo* drug	*AttributeTo*: {attributed to, secondary to, related to, …}	hypoglycaemia sec to glimepiride
3	*AllergyTo* drug	*AllergyTo*: {da to, allergic to, drug allergy, …}	Allergic to antibiotics
4	{drug *StopAfter* or *StopBefore* drug} + *word_d is*(*drug, AE*) < 12	*StopAfter*: {stop, held off, interrupt, discontinue, take off, …}; *StopBefore*: {stop, discontinue, take off, switch, change, not to start, …}	simva was discontinued after leg muscles became painful; Patient taken off vanco as renal function declined
5	{*AEBefore* AE or AE *AEAfter*} + *word_d is* (*drug, AE*) < 12	*AEBefore*: {in view of, following, noted, develop, likely to, complain of, …}; *AEAfter*: {develop, associate, …}	Complains of drowsiness after taking decol syrup; amitriptyline changed to Sidelium on discharge in view of slightly prolonged qtc 474 ms

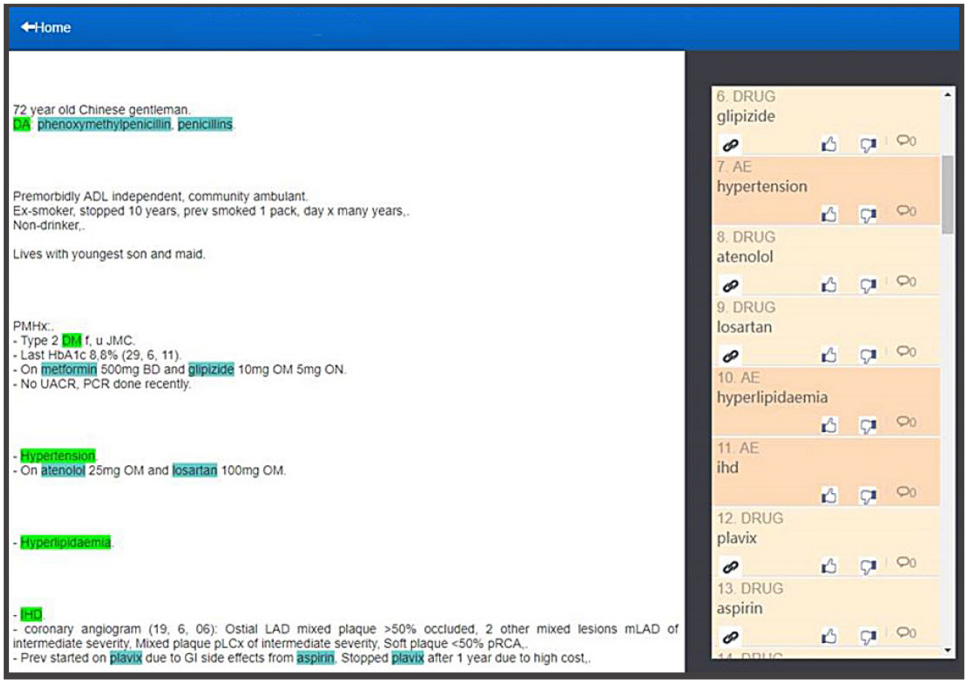

Fig. 3. Screenshot of Readpeer system annotation page. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

corresponding drug name and AE name will blink in the discharge summary. Experts can review and verify the accuracy of a machine tagged drug, AE and drug-AE pairs by voting on each tagged entity as Correct/Incorrect. If experts notice missed drug or AE names, they can highlight the missed drug or AE name and add it to the list of drugs or AEs. Importantly, if a drug-AE relation had been missed by the machine, experts can click on the drug and link it to the AE to create a new drug-AE relation annotation. Comments can also be incorporated into each annotation on the right panel.

2.4. Expert validation

Given that we adopted a rule-based approach, REAP does not require labelled data in advance. Instead, after machine extraction of drug names, AE names and drug-AE relations, all results were reviewed by human experts. If a drug term had been missed by the machine, the annotators corrected this by adding it in. If an AE term had been missed by the machine, the annotators would not add it in unless it was part of a true drug-AE pair. To ensure reliability of expert annotations, each summary was evaluated by three independent annotators during the initial training phase, and inter-annotator agreement (IAA) was measured. As presented in Fig. 4, details about a disagreement was given, including the drug name, the AE name, the context and the link to the discharge summary which contained the drug-AE pair, so that annotators could discuss and resolve conflicting opinions. Having arrived at an acceptable IAA of 88.6% and a Kappa statistic of 0.755 ($p < 0.001$) following discussions and consensus attainment, all subsequent annotations were deemed to be consistent to warrant independent annotation. No further IAA assessments were performed on subsequent annotations.

Drug: ***aspirin***
AE: ***stroke***
Link: 20111000_5835617
Creator: ***Machine***
Context:
Left LL poer was 4, 5, subsequently progressed to 1, 5 on second day, Likely progression of ***stroke***, ***aspirin*** changed to clexane 40mg BD.- lower limb strength subsequently improved in the ward to 3, 5, contin
User vote result:
Annotator 1: Wrong
Annotator 2: Correct
Annotator 3: Wrong

Fig. 4. Example for inter annotator agreement.

REAP was updated after each trial was annotated and evaluated. For trials 10 to 27, 25% of the records were randomly assigned to three annotators to measure inter-annotator agreement (IAA). Specifically, 120 discharge summaries were divided into four groups of 30 summaries. Each trial consisted of 30 unique summaries and the remaining 30 summaries were repeated in each trial, resulting in 60 discharge summary records in each trial (Fig. 5). Since REAP was relatively stable at this phase of training, it was updated every 2 rounds of experiments, i.e. every 6 trials. Finally, after the system was settled based on the observations of all previous trials, it was evaluated against 6 trials of 100 discharge summaries each.

2.5. Using external dataset i2b2 to assess generalizability of REAP

In order to evaluate the generalization of our algorithm, we also tested it on 100 randomly selected electronic health reports (EHR) from i2b2 2009 Medication Challenge [19]. These are deidentified discharge summaries of Intensive Care Unit patients from the Partners Healthcare network in the United States.

3. Results

The hospital discharge summaries were obtained from patients who were discharged from NUH in 2011. A total of 1620 discharge summaries had been annotated (180 excluded for incompleteness), of which 1020 were used during the training phase of REAP while the remaining 600 were used for validation. Overall, the discharge summaries comprised of patients discharging from the following departments; internal medicine (12.1%), cardiology (11.6%), oncology (5.7%), general surgery (4.1%), geriatric medicine (2.1%). The remaining discharge summaries included several smaller sub-specialities including nephrology, rheumatology, infectious disease, dermatology, neurosurgery and urology among others.

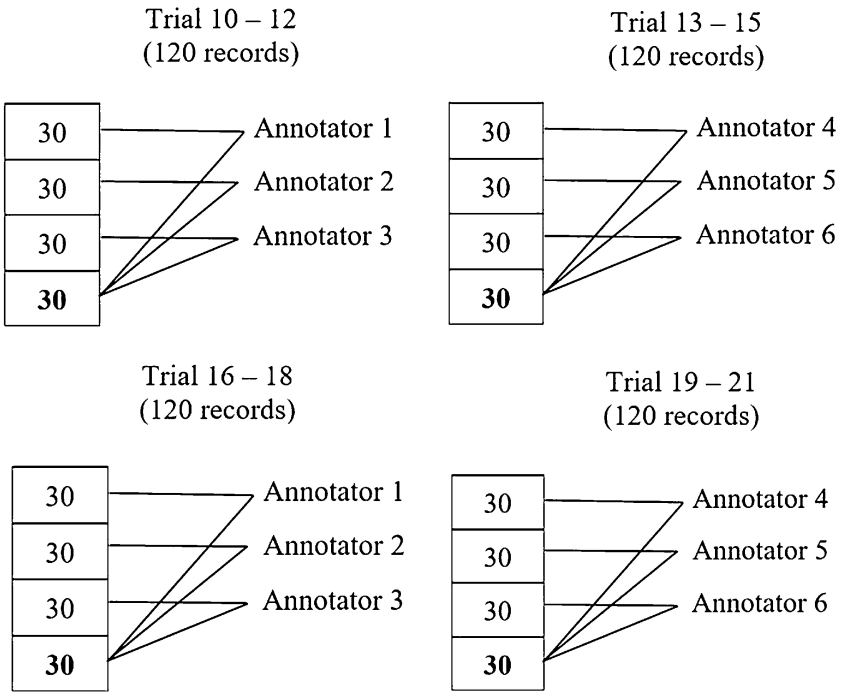

Fig. 5. Assignment of EHRs among annotators.

As shown in Table 2, we built our preliminary version of dictionaries and rules based on 300 summaries from trials 4 to 9, each consisting of 50 of discharge summaries. Trials 1 to 3 were used to allow the annotators to be familiarised with the user interface of Readpeer-HSA.

3.1. Drug and AE name recognition

Table 3 demonstrates the performance of REAP on drug and AE name recognition. Two metrics were used to measure the performance of named entity recognition, namely precision and recall. Precision refers to the ratio of correct annotations out of all results annotated by REAP while recall refers to the ratio of correct annotations recognized compared to all correct annotations. For trial 4 to 9, excellent performance was achieved for drug entity recognition, over 90% precision and recall were achieved over all trials. As the dictionaries were updated after each trial, the performance improved with each trial. REAP achieved an average precision and recall of 0.923 and 0.930 for drug name recognition on six sets of 50 manually reviewed records, and an average precision and recall of 0.949 and 0.968 for AE recognition. To note that we did not hunt for missed AEs unless there is a drug-AE relationship. After which, we decided to shift our focus to evaluating the drug-AE relations for subsequent trials.

3.2. Drug-AE relation extraction

As shown in Table 4, REAP achieved a weighted average performance of over 75% precision and about 60% recall on test data trial 28 to 33 for identification of valid drug-AE relations. From the weighted average performances of training phase, we can see that the precision keeps increasing while the recall falls, which is a common trade-off for rule-based systems.

Table 5 demonstrates the performance of REAP on i2b2 dataset. Although dealing with an unseen dataset with different writing styles, REAP was able to achieve high accuracy and recall for drug and AE name recognition. Meanwhile, both the precision and recall for detecting drug-AE relations is lower than on NUH dataset. We believe that different culture and writing styles adopted by various healthcare institutions could affect the performance of REAP and we hope to bridge the gap between datasets by introducing learning based models in future.

Table 2
Dataset summary.

Dataset	Trial no.	No. of records per trial	Total size	Purpose
NUH	4–9	50	300	Train
NUH	10–27	60	720	Train
NUH	28–33	100	600	Validation
i2b2	–	100	100	Validation

Table 3
Drug and AE name recognition results.

Trial no.	No. of records	Drug precision	Drug recall	AE precision	AE recall
4	50	0.928	0.902	0.912	0.933
5	50	0.925	0.938	0.957	0.942
6	50	0.927	0.905	0.984	0.993
7	50	0.920	0.953	0.932	0.948
8	50	0.912	0.919	0.958	0.994
9	50	0.928	0.965	0.949	0.998
	Avg	0.923	0.930	0.949	0.968

4. Discussion

In this study, we report the development and validation of REAP, an expert-derived, rule-based algorithm to detect true drug-AE pairs which are semantically connected in unstructured electronic medical records data (hospital discharge summaries). On validation of an independent set of 600 discharge summaries, REAP demonstrates a precision and

Table 4
REAP performance for drug-AE relation extraction.

Trial	No. of records	Precision	Recall	*F*-score
		(a) Training phrase		
4	50	0.340	0.580	0.429
5	50	0.520	0.684	0.591
6	50	0.818	0.794	0.806
7	50	0.727	0.500	0.593
8	50	0.704	0.760	0.731
9	50	0.737	0.609	0.667
Avg		0.641	0.655	0.648
10	120	0.676	0.694	0.685
11		0.679	0.527	0.593
12		0.719	0.767	0.742
13	120	0.800	0.610	0.692
14		0.774	0.750	0.762
15		0.500	0.538	0.518
Avg		0.701	0.644	0.671
16	120	0.423	0.727	0.535
17		0.500	0.696	0.582
18		0.447	0.809	0.576
19	120	0.846	0.660	0.742
20		0.408	0.500	0.449
21		0.425	0.404	0.414
Avg		0.507	0.597	0.548
22	120	0.667	0.333	0.445
23		0.750	0.478	0.584
24		0.682	0.484	0.566
25	120	0.826	0.605	0.699
26		1.000	0.465	0.635
27		0.913	0.651	0.760
Avg		0.784	0.530	0.632
		(b) Validation Phrase		
28	100	0.719	0.495	0.586
29	100	0.773	0.472	0.586
30	100	0.729	0.632	0.677
31	100	0.755	0.74	0.747
32	100	0.844	0.692	0.761
33	100	0.773	0.618	0.687
Avg		0.757	0.586	0.661

Table 5
Results from applying REAP on the i2b2 dataset consisting of 100 records, with 129 drug-AE associations.

Type	Precision	Recall	*F*-score
Drug	0.936	0.924	0.930
AE	0.811	0.979	0.887
Drug-AE	0.641	0.457	0.534

recall of 0.757 and 0.586, respectively. Against a backdrop of severe under-reporting of ADRs to drug surveillance agencies which delays safety signal detection, REAP carries considerable potential in picking up true drug-AE pairs that can be subsequently made known to surveillance agencies as another tool in the pharmacovigilance toolkit.

Although a recall of 0.586 may appear modest, to the best of our knowledge, there have not been any similar algorithms, tools or programmes that perform the same task as REAP does to date, which is to detect semantically linked drugs and AEs. Currently, where 80% of actual ADR occurrences do not get reported, a tool that detects 58.6% of all true drug-AE pairs in unstructured text can already provide much needed signals to national pharmacovigilance agencies to enhance drug safety monitoring [1]. It is worth noting that detecting true drug-AE pairs in unstructured text is an extremely challenging task, because clinicians can represent drug-AE pairs in a variety of different ways. Any tool designed to address this challenge will have to be able to handle medical jargon, misspellings, symbols and accommodate for the local documentation norms of clinicians in a given setting. For instance, it is common practice for orthopaedic surgeons to refer to a fracture with the symbol '#' (e.g. spinal '#'), whereas medical oncologists may use the '#' to indicate the cycle number of chemotherapy.

REAP has achieved high precision and recall for drug and AE term extraction by introducing expert knowledge through well-constructed and targeted reference dictionaries. The drug dictionary was a compendium of ingredient and brand names used in the region, common misspellings identified from having carried out a cleaning exercise on pharmacy orders. We expanded the dictionary we used in our previous publication by adding in drug names from RxNorm and the FDA Orange Book. Fuzzy logic was used to catch spelling errors and phonemes were used to improve recall. Another new feature that we introduced to control precision was to exclude fuzzy logic for 9000 common English words that could be permuted to drug names with simple transformations. Even though REAP was trained on data in Singapore, it also performed well on the i2b2 drug challenge set for drug name extraction. Other efforts in this domain have utilized the UMLS for extracting adverse event entities [20]. We chose instead to utilise dictionaries more focused on adverse drug reaction (ADR) terminology.

Previously, we used the WHO-Adverse Reaction and Medical Dictionary for Regulatory Activities (MedDRA) Terminologies. For the current dictionary, we have added AE terms from Sider SideEffect, IMI Protect Adverse Drug Reactions Database, and the drug side effect database of Drugs.com. The advantage of constructing our own compendium of adverse event terminology was the ability to easily add common abbreviations and misspellings appropriate to our local context.

We devoted considerable effort to customizing an annotation tool for the purpose of reviewing machine annotated records. Readpeer-HSA highlights the drug and AE names in different colors, allowing the reviewer to quickly vote yes or no to machine extracted entities. The tool also makes it easy to add in terms that were missed by the machine. When the machine has identified a drug-AE pair, both highlighted terms would blink, making it easy for the reviewer to focus on the particular relationship being assessed. Faster human review is important for improving NLP algorithms, because a sizable number of annotated records are required for machine learning. For example, we are now testing semantic methodologies to better infer context of true drug-AE pairs, and we will need additional sets of annotated records to measure performance. The annotation tool also made it easy to measure the distances between drug and AE in true drug-AE relationships, which allowed us to quantitatively evaluate the optimal distance for searching. Indeed, many instances extended beyond the same sentence, declining exponentially with sentence distance. A loose rule of 3 sentence and 25 word distance in general and 12 words when certain sentence patterns existed, was helpful for controlling machine run times, yet capturing most drug-AE pairs. Other previously reported algorithms that were confined to single sentences may benefit from loosening of the distance requirement [8].

REAP is not restricted to particular drugs or drug classes or to specific medical specialties, which is an important attribute to carry out drug safety monitoring by the drug regulatory authority. The drug and AE dictionaries may be updated periodically to ensure that the overall tool remains relevant for performing surveillance activities. While our current assessment of REAP has focussed on its ability to detect drug names, AE names and true drug-AE pair, we acknowledge that there may be other means of evaluating a computational tool in relation to other possible tools available [21]. These may become relevant when there are other comparable tools or algorithms made available.

Several studies conducted in different regions of the world have documented that ADRs contribute to 2 to 10% of hospitalizations [22,23]. For example, in a recently conducted survey at another general hospital in Singapore, ADRs contributed to 8.1% of hospital admissions (excluding elective surgeries and obstetric cases), and the median length of stay in the hospital for these potentially avoidable ADRs was four days, and was one day longer than those without an ADR [22]. Polypharmacy is a known risk factor associated with readmission risk and ADRs are a significant contributor to health care costs and

utilization [24,25]. The use of REAP would help us to pick up drug related AEs through mining the hospital discharge summaries.

While the use of expert-derived rules for text-based relationship detection has been explored previously, the use of these methods to address the problem of detecting ADRs in discharge summaries has not been explored and our study demonstrates potential and value for future work in this area. On a broader level, our study also demonstrates the value and need for greater collaboration between regulatory agencies and academia to advance regulatory science and improve patient safety.

5. Conclusion

We developed an algorithm to automatically extract drug and AE names from free text hospital discharge summaries and propose a novel rule-based method to identify relations between drug and AE pairs. The performance of our rule-based method had over 75% precision and about 60% recall for drug-AE relations. Readpeer-HSA is a useful tool to facilitate reviewers to easily vote and make corrections to machine annotated results, but further validation in yet larger, external datasets are warranted to fully evaluate its potential for detecting drug-AE pairs in as part of routine pharmacovigilance.

6. Summary table

6.1. Related work on this topic

- There have been several studies focusing on extracting drug-AE relations from unstructured electric health records. Their approaches can be divided into three main categories, namely statistics-based method, keyword-based method and learning-based method.
- Statistics-based methods calculate the co-occurrence of a pair of drug and AE based on a large corpora of medical reports and then predict potential drug-AE relations from an overall view. They do not aim to detect drug-AE relations in each electric health record.
- Keyword-based methods identify a collection of trigger phrases which may indicate the presence of drug-AE relations and then employ customized rules on the syntactic and semantic patterns surrounding the drugs and AE to further determine the relation. However, existing studies usually limit the distance between drugs and AEs to be within one sentence.
- There are also learning-based methods using classifiers, such as Naive Bayes and decision tree, to classify drug-AE relations. They usually require a large amount of annotated data to train the model.

6.2. Our main contributions

- We developed an algorithm REAP that automatically extracts drug and AE terms from free-text hospital discharge summaries and have proposed a novel rule-based method to identify relations between drugs and AEs. To support the algorithm, we also built customized drug and AE gazetteers and a list of other useful gazetteers.
- We annotated 1620 free-text hospital discharge summaries from NUH for drug and AE terms and drug-AE relations and used them to train and evaluate REAP. Satisfactory precision and recall has been achieved to detect drug-AE relations.
- We implemented a customized online reading and annotation system called Readpeer-HSA to correct and vote on machine annotated results.

Conflicts of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Acknowledgement

The authors are grateful to Mr Liu Qi for assisting in the development of programming codes while he was with the Department of Computer Science, School of Computing, National University of Singapore. We appreciate Ms Celine Loke's assistance for providing valuable insights. We are grateful to the Academic Informatics Office, National University Health System for providing the data to enable HSA to conduct this study. We would also like to thank A/Prof Cheng Leng Chan, Group Director of the Health Products Regulation Group (HPRG), Dr. Dorothy Toh, Assistant Group Director, HPRG Post-market cluster and Ms Jalene Poh, Director, Vigilance and Compliance Branch at the Health Sciences Authority for their support to pursue this research to advance the organization's pharmacovigilance mission as well as Michael Winther for his valuable programmatic coordination for the SAPhIRE (Surveillance and Pharmacogenomics Initiative for Adverse Drug Reactions) Project. This study was conducted under the SAPhIRE Project, funded by a Strategic Positioning Fund grant from the Biomedical Research Council of the Agency for Science, Technology and Research of Singapore (SPF2014/001).

Appendix A. List of trigger phrases

adverse to
after starting
after taking
after
allergic
allergies
allergy
associate
associated
attribute to
attributed to
cause
caused by
caused
cessation of
change to
changed to
controlled with
converted to
da to
develop from
developed from
developed
develops
discontinue
discontinued
drug allergic
drug allergy
drug induced
drug-induced
due to
due
following
held off in view of
held off
hypersensitivity
improved with
increasing dose
induced
interrupt

likely
not continued
not to start
post
reduce
reduced
related
sec to
secondary to
secondary
side effect
stop
stopped
stopping
subsequently developed
switch
switch to
switches to
switched
switched to
take off
taken off
took off
treated with
2 to
-induced
− >
< −

Appendix B. Rules for human annotators

B.1 Rules to annotate drugs

- Annotate all drugs, even if negated. For instance, in these situations e.g. “not taking abx prescribed” and “darbepoietin stopped”, the drugs “abx” and “darbepoietin” will be annotated.
- Search for missed instances of western drugs and its abbreviations (e.g. OHGAs for oral hypoglycaemic agents), annotate it and request addition to the drug dictionary.
- Machine-annotated instances of complementary health products, supplements and electrolytes (e.g. vitamins, dextrose, KCl) will be accepted. However, missed instances need not be annotated.

B.2 Rules to annotate AEs

- If there is no drug, the AEs in discharge summaries will not be annotated.
- Apply negation for AEs.
- Where an AE is wrongly annotated by the machine, this will be tagged as incorrect, and where possible, a rule will be added to improve the algorithm. For instance, in “SpO2 100% RA”, machine annotates RA as it stands for “rheumatoid arthritis” in the gazetteer. However, in this context, RA refers to “room air”, and should not be annotated. In this scenario, the annotation is marked as incorrect, with a rule added to look for contextual clues (e.g. % before RA) for negation.
- Missed AEs need not be annotated, unless here is a drug-AE relationship.

B.3 Rules to annotate drug-AE relationships

- All drug-AE relationships will be annotated. Repeated mentions of either drugs or AEs in a single relationship only needs to be annotated once. For instance, in “GCSF related symptoms/Patient had jaw pain related to GCSF administration”, only one instance of the relationship needs to be annotated, even though GCSF is mentioned twice. If such a relationship is annotated more than once by the machine, it would still be considered correct.

References

[1] E. Lopez-Gonzalez, M.T. Herdeiro, A. Figueiras, Determinants of under-reporting of adverse drug reactions, Drug Saf. 32 (1) (2009) 19–31.
[2] R. Eriksson, P.B. Jensen, S. Frankild, L.J. Jensen, S. Brunak, Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text, J. Am. Med. Inform. Assoc. 20 (5) (2013) 947–953.
[3] S.V. Iyer, R. Harpaz, P. LePendu, A. Bauer-Mehren, N.H. Shah, Mining clinical text for signals of adverse drug-drug interactions, J. Am. Med. Inform. Assoc. 21 (2) (2013) 353–362.
[4] J. Ballard, M. Rosenman, M. Weiner, Harnessing a health information exchange to identify surgical device adverse events for urogynecologic mesh, AMIA Annual Symposium proceedings, vol. 2012 American Medical Informatics Association (2012) p1109.
[5] C. Ferrajolo, P.M. Coloma, K.M. Verhamme, M.J. Schuemie, S. de Bie, R. Gini, R. Herings, G. Mazzaglia, G. Picelli, C. Giaquinto, et al., Signal detection of potentially drug-induced acute liver injury in children using a multi-country healthcare database network, Drug Saf. 37 (2) (2014) 99–108.
[6] J. Pathak, R.C. Kiefer, C.G. Chute, Using linked data for mining drug-drug interactions in electronic health records, Stud. Health Technol. Inform. 192 (2013) 682.
[7] P.M. Kilbridge, L.A. Noirot, R.M. Reichley, K.M. Berchelmann, C. Schneider, K.M. Heard, M. Nelson, T.C. Bailey, Computerized surveillance for adverse drug events in a pediatric hospital, J. Am. Med. Inform. Assoc. 16 (5) (2009) 607–612.
[8] Y. Miura, E. Aramaki, T. Ohkuma, M. Tonoike, D. Sugihara, H. Masuichi, K. Ohe, Adverse-effect relations extraction from massive clinical records, Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2010) (2010) 75–83.
[9] S. Visweswaran, P. Hanbury, M. Saul, G.F. Cooper, Detecting adverse drug events in discharge summaries using variations on the simple bayes model, AMIA Annual Symposium Proceedings, vol. 2003 American Medical Informatics Association (2003) 689.
[10] S. Sohn, J.-P.A. Kocher, C.G. Chute, G.K. Savova, Drug side effect extraction from clinical narratives of psychiatry and psychology patients, J. Am. Med. Inform. Assoc. 18 (Supplement_1) (2011) i144–i149.
[11] T. Munkhdalai, F. Liu, H. Yu, Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning, JMIR Public Health Surveill. 4 (2) (2018).
[12] S. Santiso, A. Perez, A. Casillas, Exploring joint ab-lstm with embedded lemmas for adverse drug reaction discovery, IEEE J. Biomed. Health Inform. (2018).
[13] M. Oronoz, K. Gojenola, A. Pérez, A.D. de Ilarraza, A. Casillas, On the creation of a clinical gold standard corpus in Spanish: mining adverse drug reactions, J. Biomed. Inform. 56 (2015) 318–332.
[14] F. Liu, A. Jagannatha, H. Yu, Towards drug safety surveillance and pharmacovigilance: current progress in detecting medication and adverse drug events from electronic health records, Drug Saf. (2019).
[15] P. San Ang, L.Y. Fan, M.Y. Tham, S.H. Tan, S.B. Soh, B.P. Foo, C.W. Loke, S. Hu, C. Sung, Towards human-machine collaboration in creating an evaluation corpus for adverse drug events in discharge summaries of electronic medical records, Big Data Res. 4 (2016) 37–43.
[16] A.R. Aronson, Metamap:, Mapping Text to the UMLS Metathesaurus, NLM, NIH, DHHS, Bethesda, MD, 2006, pp. 1–26.
[17] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, C.G. Chute, Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications, J. Am. Med. Inform. Assoc. 17 (5) (2010) 507–513.
[18] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, B.G. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries, J. Biomed. Inform. 34 (5) (2001) 301–310.
[19] Ö. Uzuner, I. Solti, F. Xia, E. Cadag, Community annotation experiment for ground truth generation for the i2b2 medication challenge, J. Am. Med. Inform. Assoc. 17 (5) (2010) 519–523.
[20] X. Wang, H. Chase, M. Markatou, G. Hripcsak, C. Friedman, Selecting information in electronic health records for knowledge acquisition, J. Biomed. Inform. 43 (4) (2010) 595–601.
[21] S. Siadat, A.M. Ferreira, T.T. Khoei, A.R. Ghapanchi, Performance analysis of qos-based web service selection through integer programming, World Appl. Sci. J. 28 (4) (2013) 463–472.
[22] S.L. Chan, X. Ang, L.L. Sani, H.Y. Ng, M.D. Winther, J.J. Liu, L.R. Brunham, A. Chan, Prevalence and characteristics of adverse drug reactions at admission to hospital: a prospective observational study, Br. J. Clin. Pharmacol. 82 (6) (2016) 1636–1646.
[23] A. Chan, D. Soh, Y. Ko, Y.-C. Huang, J. Chiang, Characteristics of unplanned hospital admissions due to drug-related problems in cancer patients, Support. Care Cancer 22 (7) (2014) 1875–1881.
[24] S.R. Dorajoo, V. See, C.T. Chan, J.Z. Tan, D.S. Tan, S.M. Abdul Razak, T.T. Ong, N. Koomanan, C.W. Yap, A. Chan, Identifying Potentially Avoidable Readmissions: A Medication-Based 15-Day Readmission Risk Stratification Algorithm, Pharmacotherapy 37 (2017) 268–277.
[25] S.L. Chan, H.Y. Ng, C. Sung, A. Chan, M.D. Winther, L.R. Brunham, H.L. Wee, Economic burden of adverse drug reactions and potential for pharmacogenomic testing in Singaporean adults, Pharmacogenomics J (2018).